Sample Sizes Required To Achieve Desired Precision Goals Within A Stratum.

Marianna Alexandersdottir
NWIFC

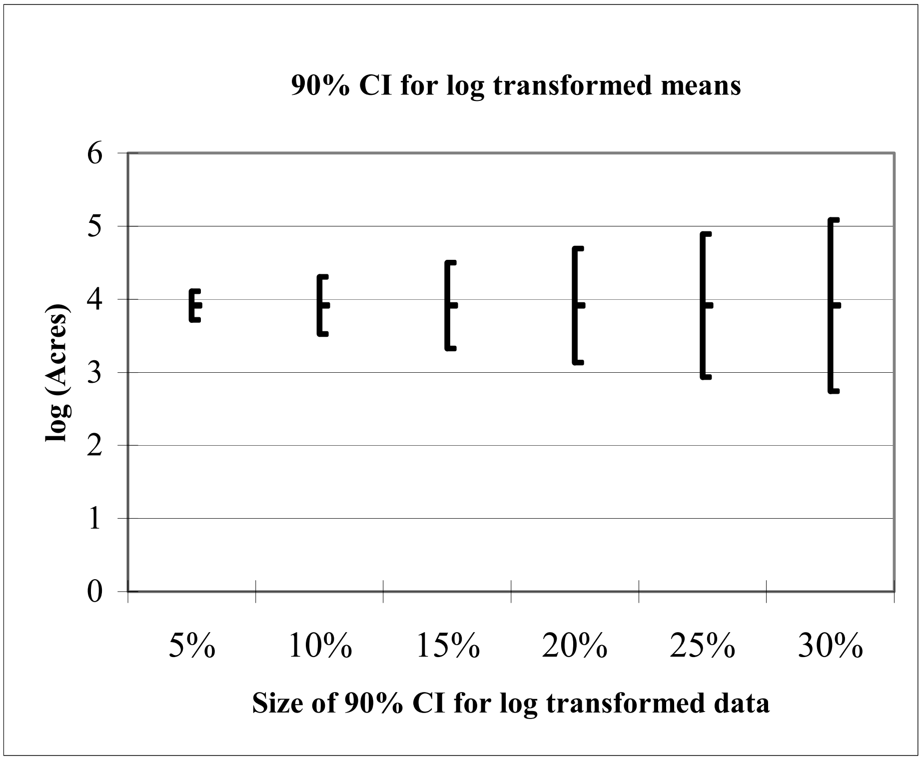
Evaluation of sample sizes required to achieve project objectives is only one part of a sample design. But assuming random sampling, the data gathered already can be used to provide some guidelines as to sample size requirements within a stratum. In this evaluation two strata are discussed, a stratum for the 52 acre default area region and a stratum within the 300 acre default area region. These strata could be ecoregions, the entire default area region or some other geographically defined areas.

The data analysis suggested that the basin area measurements have a distribution skewed to the left, with a longer right tail. A log-transformation appears to normalize the data and the median is a better estimator of the center of the distribution than the mean of the basin areas. The mean of the log-transformed data provides an estimate of the median when transformed back to the original scale.

The estimation of the mean of the log-transformed data with a 90% confidence interval (CI) also provides an estimated confidence interval on the original scale by transforming the mean and lower and upper bounds back to the original scale. The 90% CI of the log-transformed data will be symmetrical (the distance to the lower and upper bounds are equal), but when transformed back to the original scale the 90% CI will no longer be symmetrical, but will be wider on the upper end of the range (Figure 1, Table 1). In addition the size of the confidence interval on the original scale is larger that at the log-transformed scale, where the relative size is the distance to the lower or upper bound divided by the estimated median. A 90% CI with a relative size of ±10% for the log-transformed mean will translate to a confidence interval that ranges 32% of the estimated median to the lower bound and 48% to the upper bound (Table 1).

Table 1. Relative size of 90% confidence intervals at the log-transformed and original scales.

Relative Size of 90% CI for log-transformed data	Original Scale		
	Estimated median	Size for lower bound	Size for upper bound
5%	50	18%	22%
10%	50	32%	48%
15%	50	44%	80%
20%	50	54%	119%
25%	50	62%	166%
5%	300	25%	33%
10%	300	43%	77%
15%	300	57%	135%
20%	300	68%	213%
25%	300	76%	316%



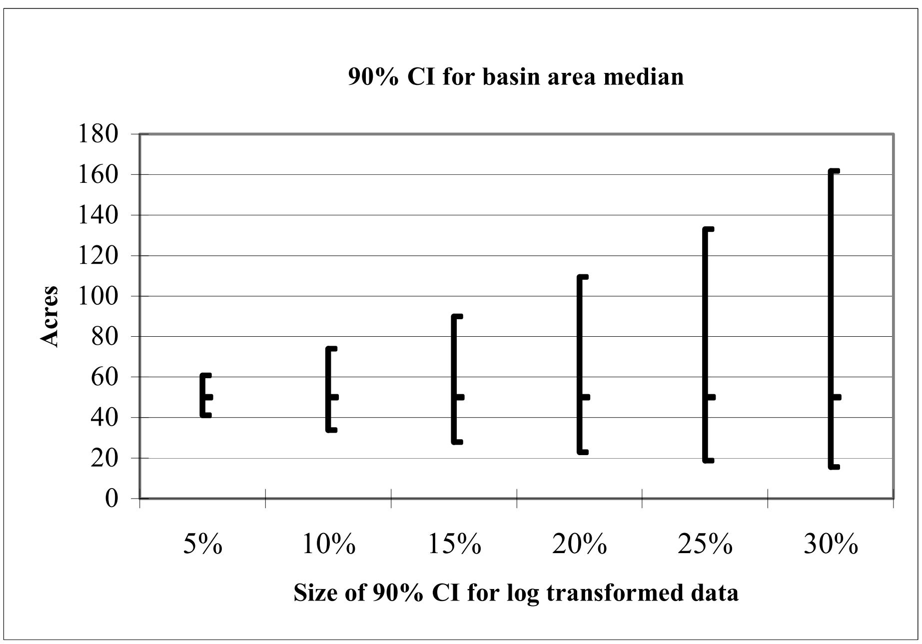


Figure 1. **90% Confidence intervals for log-transformed means and median in original scale relative to the size of the 90% CI of the log transformed mean. The CI for the log-transformed mean are symmetrical (upper graph), but the CI are not symmetrical when transformed back to the original scale (lower graph).**

1 Evaluation of Sample Sizes

The approximate 90% confidence interval for the log-transformed mean is estimated using a normal Z-statistic by:

$$\boldsymbol{Mean} \pm \frac{\boldsymbol{Standard\ Deviation}}{\sqrt{\boldsymbol{n}}} \bullet 1.65$$

This provides a method to estimate sample sizes needed to achieve desired precision levels defined by the relative size of the confidence interval by;

$$\boldsymbol{n} = \boldsymbol{CV}^2 \frac{1.65^2}{\boldsymbol{r}^2}$$

where $\boldsymbol{r}$ is the relative size of the confidence interval (i.e. $\left(\frac{\boldsymbol{Standard\ Deviation}}{\sqrt{\boldsymbol{n}}} \bullet 1.65\right) = \boldsymbol{r} * \boldsymbol{Mean}\}$

and $\boldsymbol{CV}$ is the coefficient of variation of the population ($\frac{\boldsymbol{Standard\ Deviation}}{Mean} * 100$).

1.1 Sample sizes

In order to evaluate sample sizes required to achieve the desired precision, three pieces of information are required:

1. The expected estimate of basin area. Here, we have used 52 and 300 acres, default basin areas used for the Westside and Eastside regions.
2. The expected variance, or CV, of the basin areas. Here, we have used the CVs estimated in sampling already accomplished (Table 2).
3. A measure of desired precision for the "new" estimate. Here, we have used the relative size of the 90% confidence interval as a measure of the precision.

A three step procedure is used:

1. It is necessary to decide what precision is desired for the estimates on the original scale, i.e., the median of the measured basin areas.
2. Once this is decided, the precision (or size of 90% CI) required for the log-transformed mean estimation can be determined using the graphs in Table 1.
3. Finally, given the expected basin areas (52 and 300 acres) and the expected CVs then Table 3 can be used to find the sample size required to achieve the desired precision, given the estimate and variances expected.

Table 2. Estimated Mean, Standard Deviation , coefficient of variation (CV)[1] and number of sites sampled for log transformed data by cooperator, ecoregion and default area stratum.

		log Pp				log Pd			
Ecoregion	Coop	Mean	Std. Dev.	CV	N	Mean	Std. Dev.	CV	N
300 Acre Stratum									
15	COL	5.5961	1.0835	19.4%	6	5.6605	1.2727	22.5%	5
15	SPO	4.9777	2.2261	44.7%	5	4.2188	2.1351	50.6%	6
Ecoregion average				32.0%				36.5%	
15 combined		5.3150	1.6351	30.8%	11	4.8742	1.8692	38.3%	11
77 East	COL	3.6600	0.8538	23.3%	5	3.6687	0.7628	20.8%	6
77 East	LVF	3.8102	1.7107	44.9%	11	2.6510	1.7949	67.7%	12
Ecoregion average				34.1%				44.2%	
77 E combined		3.7632	1.4665	39.0%	16	2.9902	1.5809	52.9%	18
9	LVF	4.1682			1	3.7466	0.5962		2
Stratum average				34.6%				45.2%	
300 Acre Stratum combined		4.3873	1.6638	37.9%	28	3.7075	1.8441	49.7%	31
50 Acre Stratum									
1	DFW	3.1447	1.1012	35.0%	18	1.8641	0.7685	41.2%	23
1	HOH	1.1742	1.6005	136.3%	17	0.8148	1.5426	189.3%	19
1	LVF	1.6541	0.6874	35.0%	3	1.8361	0.5177	28.2%	3
Ecoregion average				68.8%				86.2%	
1 combined		2.1700	1.6161	74.5%	38	1.4192	1.2466	87.8%	45
2	LVF	5.1472	2.2063	42.9%	4	2.9642	1.7260	58.2%	4
2	PGS	3.0350			1	2.7850			1
2	SUQ	2.4366	1.1922	48.9%	6	2.1772	1.2031	55.3%	6
Ecoregion average				45.9%				56.7%	
2 combined		3.4767	1.9889	57.2%	11	2.5186	1.3319	52.9%	11
4	LVF	2.3716	1.0344	43.6%	17	2.0214	1.1143	55.1%	17
4	TCG	2.8268	1.5526	54.9%	50	2.5991	1.5424	59.3%	57
4	YAK	3.7205	1.2394	33.3%	13	2.4093	1.4034	58.2%	13
Ecoregion average				44.0%				57.6%	
4 combined		2.8753	1.4556	50.6%	80	2.4578	1.4512	59.0%	87
77 West	SSC	2.2985	1.0290	44.8%	23	1.2832	0.7368	57.4%	18
Stratum average				56.8%				64.3%	
50 Acre Stratum combined		2.6552	1.5226	57.3%	152	2.0404	1.4182	69.5%	161

[1] $CV = \frac{\textbf{Standard Deviation}}{Mean} * \mathbf{100}$

1.1.1 An example.

We are using 50 and 300 acres as our expected estimate and Table 2 shows the CVs estimated for the samples taken in the previous sample. The pooled CV for log Pd in the 300-acre default is about 50% and for the 52-acre default, it is about 70%.

It is desired to have a 90% confidence interval that does not range more than 50% on either side for an estimated median of 50 acres and 300 acres.

Looking at Table 1 we would require a 90% confidence interval for the log-transformed data of no larger than ±10% for the 50 acre estimate and 5% for the 300 acre estimate to achieve this precision goal.

Using a relative size of the 90% CI of ±10% for the log-transformed mean, and a CV of 50% a sample size per stratum of 68 is required and a CV of 70% results in a sample size requirement of 133 (Table 3). With these sample sizes Table 1 indicates that the median for the 300-acre default will be about +77%, -43% and for the 52-acre default it will +48%, - 32%. If this is not sufficient for the 300 acre region, then sample sizes would have to be increased to 534 sites to achieve a 90% CI of ±5% for the log-transformed mean, which would translate to a 90% CI ranging from –25% to +33% around an estimate of 300 acres (Table 1).

Table 3. Sample sizes required to estimate 90% CI of mean of log-transformed basin areas for a given relative size and CV of the transformed data.

CV of log-transformed mean Basin Area	Relative Size of 90% CI of log-transformed mean Basin Area				
	5%	10%	15%	20%	25%
5%	3	1	0	0	0
10%	11	3	1	1	0
15%	25	6	3	2	1
20%	44	11	5	3	2
25%	68	17	8	4	3
30%	98	25	11	6	4
35%	133	33	15	8	5
40%	174	44	19	11	7
45%	221	55	25	14	9
50%	*272*	*68*	*30*	*17*	*11*
55%	329	82	37	21	13
60%	392	98	44	25	16
65%	460	115	51	29	18
70%	*534*	*133*	*59*	*33*	*21*
75%	613	153	68	38	25
80%	697	174	77	44	28
85%	787	197	87	49	31
90%	882	221	98	55	35
95%	983	246	109	61	39
100%	1,089	272	121	68	44